Computers and Electrical Engineering 000 (2016) 1–10

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Dynamic replication to reduce access latency based on fuzzy logic system[☆]

Tao Wang[a,b], Shihong Yao[c], Zhengquan Xu[a,b,*], Shaoming Pan[a]

[a] Collaborative Innovation Center for Geospatial Technology, Wuhan University, Luoyu Road 129, Wuhan, China
[b] State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Luoyu Road 129,Wuhan, China
[c] Faculty of information engineering, China University of Geosciences, Lumo Road 388, Wuhan, China

ARTICLE INFO

Article history:
Received 5 July 2016
Revised 21 November 2016
Accepted 21 November 2016
Available online xxx

Keywords:
Data replication
Access latency
Fuzzy logic system
Simulations

ABSTRACT

In a distributed environment, limited available bandwidth resources lead a high data access latency. Replication is a popular method that can upgrade the access performance and increase the data availability. However, unreasonable replication would cause overconsumption of system resources and finally a further deterioration on data access latency. So, In this paper, a theoretical model of access latency optimization with replication is presented firstly, which complement the blank space, and then a well-designed dynamic replication strategy is proposed, which consists of three algorithms: replica selection algorithm, replica layout algorithm and replica replacement algorithm. Replica selection algorithm selects the optimal replica with a hierarchical time cost based on the derivation of the theoretical model. Replica layout algorithm selects the optimal node for placing the replica based on the spatio-temporal locality of data access. Replica replacement algorithm, in which the fuzzy logic system is introduced originally, deletes replica when the available storage space is insufficient. FLSDR is tested by OptorSim and experimental results show that FLSDR achieves better performance in comparison with other algorithms in terms of mean job execution time, computing resource usage, number of data scheduling between clusters and number of replicas.

1. Introduction

Access latency is a major concern of users which reflects data access performance for data intensive applications. In such applications, massive datasets are processed and generated worldwide. Hence, how to optimize the access latency widely deserves the attention of investigators and engineers. Prefetching, cache and data replication are the three popular methods used to improve the access latency. In a distributed environment, access latency primarily depends on the availability of network bandwidth, because transferring large-sized files usually takes much bandwidth for a long time [1]. Furthermore, data availability always restrict the performance of access latency, and if the datasets required by a job are unavailable, the job will be hung up until a user terminate it. A consensus shows that replication is the only way to improve the data availability in distributed data storage [2]. Consequently, replication has gradually become a hot topic in the field.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. Z. Zheng.
* Corresponding author.
E-mail addresses: wangtao.mac@whu.edu.cn (T. Wang), xuzq@whu.edu.cn (Z. Xu).

http://dx.doi.org/10.1016/j.compeleceng.2016.11.022
0045-7906/© 2016 Published by Elsevier Ltd.

As each storage node in a distributed system has limited storage capacity and the data volume continues to grow, the data replication is facing more problems [3], and replica selection, replica layout and replica replacement are faced with higher requirements, particularly replica replacement. Some previous works on replication are in an extreme that is pure theoretical exploration based on many theoretical hypothesis, such as infinite storage capacity of each storage node, divorced from practice. These works can be of hardly any usefulness to industrial systems. In contrast, most of the other works are in the other extreme that is heuristics-based experiments without theoretical analysis. These works are meaningful to some certain kinds of application environment, however not universally applicable and portable.

In this paper, we devote to minimize the access latency by replicating the datasets into storage nodes with limited storage space, for storage nodes to finish executing their jobs. Specifically, the main contributions of this paper are as follows: (1) we originally formulate the model of access latency optimization, which provably gives quantitative analysis on the replications and the data access latency reduction and can expanded to any distributed storage system, rectifying the lack of theoretical analysis; (2) we propose a novel dynamic replication strategy named FLSDR which contains replica selection, replica layout and replica replacement. On this basis of the quantitative analysis, FLSDR always selects the optimal replica based on a hierarchical transferring time of data scheduling, and places replicas onto the optimal node based on the spatio-temporal locality of data access; (3) we originally introduce the fuzzy logic system into the replica replacement algorithm of dynamic replication strategy. The algorithm defines and calculates the replica life value (RLV) of the replicas and handily deletes the replica with the minimum RLV when the available storage space is insufficient. Consequently, large amounts of experimental results demonstrate that mean job executing time, data access latency and bandwidth consumption of FLSDR decrease significantly for any job scheduling strategy comparing other dynamic replication strategies.

The rest of the paper is organized as follows. Section 2 gives an overview of previous work on data replication. The model of access latency optimization is described in Section 3. Section 4 presents the proposed FLSDR. The simulation results are introduced and analyzed in Section 5. Finally, Section 6 concludes the paper and discusses our future work.

2. Related works

Currently, most of studies on replication strategies are concentrated on the domain of data grid and many investigators have valuable achievements. At the very beginning, replication strategies are static and the created replica will exist in the same place until the user deletes it manually or its duration is expired. These strategies cannot adapt to the changes of accessing behaviors of large scale users and system state. Then, dynamic replication strategies, which can accommodate various changes and create, delete and manage replicas automatically in stochastically fluctuating distributed data storage, has become a critical research issue for investigators. Hence, for the purpose of this paper, only the dynamic strategy is given consideration.

Carman et al. [4] and Bell et al. [5] proposed dynamic replication strategies based on economic models. Tang et al. [6] presented Simple Bottom-Up (SBU) and Aggregate Bottom-Up (ABU) for multi-tier grid to reduce the data access latency. Based on temporal locality, Chang et al. [7] proposed the Latest Access Largest Weight (LALW) to find out a popular file for replication. The studies are node-level based. However, the most critical factor impacting data access performance is bandwidth availability, which should be mainly considered. Thus, Bandwidth Hierarchy Based Replication (BHR) [8], the modified BHR [9] and the Dynamic Hierarchy Replication (DHR) [10] are proposed successively. However, their performance degrade sharply when the number of nodes increase. Meanwhile, some studies on the combination of job scheduling strategies and dynamic replication strategies are proposed, in which the Integrated Replication and Scheduling Strategy (IRS) [11], the Hierarchical Cluster Scheduling Strategy (HCS) and the Hierarchical Replication Strategy (HRS) [12] and the Weighted Scheduling Strategy (WSS) and Enhanced DHR (EDHR) [13] are the outstanding results. However, job scheduling is impacted potentially by the characteristics of the job, and when the characteristics and the requirements of jobs change, the established replication strategy is no longer valid.

Whats more, because of the limited storage resources and the variation of data access, replication replacement algorithm is of great concern on keeping the replication strategy long-term validity. But, the dissatisfactory is that the rough the Least Recently Used (LRU) and the Least Frequently Used (LFU) are frequently used to replace replicas and unaccommodated in distributed environment. The economic models based replication replacement algorithm [13] deletes the replicas in batches, with low accuracy. Hence, the problem formulations of dynamic replication strategy are as follows:

(1) The existing works, heuristics-based without theoretical analysis, are meaningful to some certain kinds of application environment, however not universally applicable and portable. (2) The existing dynamic replication strategies are tight coupling to the job scheduling strategies, and once the characteristics and the requirements of jobs change, the performance and availability of dynamic replication strategies cannot be guaranteed. (3) The existing replication replacement algorithms cannot accurately delete the replicas that should be deleted with a low complexity and recreate the deleted replicas at a fraction of the time cost.

A well-designed dynamic replication strategy is require to be always available independently of job scheduling strategy, always optimize replica creation, replica layout and replica replacement from the view of global access management and suitable for large-scale distributed systems. Unfortunately, the existing dynamic replication strategies might not fit the requirements.